REVIEW ARTICLE

OPEN

Critical Data for Critical Care: A Primer on Leveraging Electronic Health Record Data for Research From Society of Critical Care Medicine's Panel on Data Sharing and Harmonization

ABSTRACT: A growing body of critical care research draws on real-world data from electronic health records (EHRs). The bedside clinician has myriad data sources to aid in clinical decision-making, but the lack of data sharing and harmonization standards leaves much of this data out of reach for multi-institution critical care research. The Society of Critical Care Medicine (SCCM) Discovery Data Science Campaign convened a panel of critical care and data science experts to explore and document unique advantages and opportunities for leveraging EHR data in critical care research. This article reviews and illustrates six organizing topics (data domains and common data elements; data harmonization; data quality; data interoperability and digital infrastructure; data access, sharing, and governance; and ethics and equity) as a data science primer for critical care researchers, laying a foundation for future publications from the SCCM Discovery Data Harmonization and Sharing Guiding Principles Panel.

KEYWORDS: common data elements; data governance; data quality; electronic health record; interoperability

Smith F. Heavner, PhD, RN[1,2]

Vishakha K. Kumar, MD, MBA[3]

Wes Anderson, PhD[1]

Tamara Al-Hakim, MD[3]

Pam Dasher, MLS[1]

Donna Lee Armaignac, PhD, APRN, CCNS, CCRN[4]

Gilles Clermont, MD, MS[5,6]

J. Perren Cobb, MD, FACS, FCCM[7,8]

Sean Manion, PhD[9]

Kenneth E. Remy, MD, MHSc, MSCI, FCCM[10,11]

Karin Reuter-Rice, PhD, NP, FAAN, FCCM[12,13]

Melissa Haendel, PhD, FACMI[14]

Society of Critical Care Medicine (SCCM) Discovery Panel on Data Sharing and Harmonization

Patients in the ICU generate massive quantities of data with "high throughput" data sources (e.g., ventilators, arterial lines) generating a continuous stream of measurements and waveforms, often represented differently across healthcare and electronic health record (EHR) systems. The need for rapid decision-making combined with the technological advancements in data science within healthcare are creating new opportunities, as well as challenges, for clinicians and researchers (1–3). Effective and efficient analysis of EHR data requires foundational knowledge and understanding of the application of key principles governing data implementation (4). Broad frameworks for data management and governance exist and should be referenced for general guidance, especially the Findability, Accessibility, Interoperability, and Reusability guiding principles (5).

Many clinicians are familiar with EHR interfaces, which allow them to view information about their patients' care at other health systems (6, 7). Ideally, these interfaces would be real-time, ubiquitous, and accessible across the continuum of care (8), but more than half of U.S. hospitals report limitations in exchanging data to support clinical care due to limitations of EHR platforms (9). These challenges are not limited to clinical care and may, in fact, be exacerbated when sharing data for research (7, 10, 11). Furthermore, patients are supportive of their healthcare data being used to improve care and understanding of clinical processes (12), provided such use addresses specific, and justifiable, concerns about privacy and oversight (13).

DOI: 10.1097/CCE.0000000000001179

The Society of Critical Care Medicine (SCCM) Discovery has launched the Data Science Campaign (DSC), an ambitious initiative aimed at improving outcomes for critically ill patients through the strategic use of large-scale data. At the heart of the DSC is the belief that data, when properly harnessed, has the potential to revolutionize critical care. Discovery has convened the Panel on Data Sharing and Harmonization (PDSH) to explore available evidence and curate guidance for the extraction, use, and sharing of patient-level data (**Fig. 1**). In this article, we present and discuss six topics identified and defined through the DSC as foundational to the generation of guiding principles, along with consensus definitions from the PDSH leads and selected examples of existing data management and use principles and tools. Here, we provide the context and framework that fosters understanding of data science principles for critical care clinicians and researchers as a primer for subsequent publications from the DSC.

METHODS

Throughout a series of DSC meetings, both in-person and virtual, this formative evaluation sought to capture, analyze, and report perspectives of participating subject matter expert (SME) on data science in critical care practice and research in a conceptual framework as part of a modified Delphi (14–16). The DSC convened a panel of SMEs representing critical care clinicians, data scientists, ethicists, industry representatives, and patient advocates. Supported by a Delphi methodologist and librarians, the PDSH deliberated use cases, unmet needs, and opportunities for data science throughout critical care practice and research. Participant perspectives were captured through meeting recordings, field notes, and unstructured interviews. Rapid evaluation and assessment methodologies (REAM) were applied to identify themes, topics, and core concepts from stakeholder perspectives. REAM provides a systematic approach to balancing efficiency and validity when time constraints are present (17, 18).

SME with specific expertise in clinical informatics and advanced analytics (including causal inference and machine learning), as well as insight into the operational goals of the DSC, conducted independent analyses of stakeholder perspectives. Findings were combined into a draft conceptual framework, which included thematic topics and definitions. The framework was refined through iterative rounds of member checking with the PDSH to improve validity and representativeness (19).

The PDSH unanimously adopted and leveraged the framework to conduct a modified Delphi seeking to identify guiding principles for data sharing and

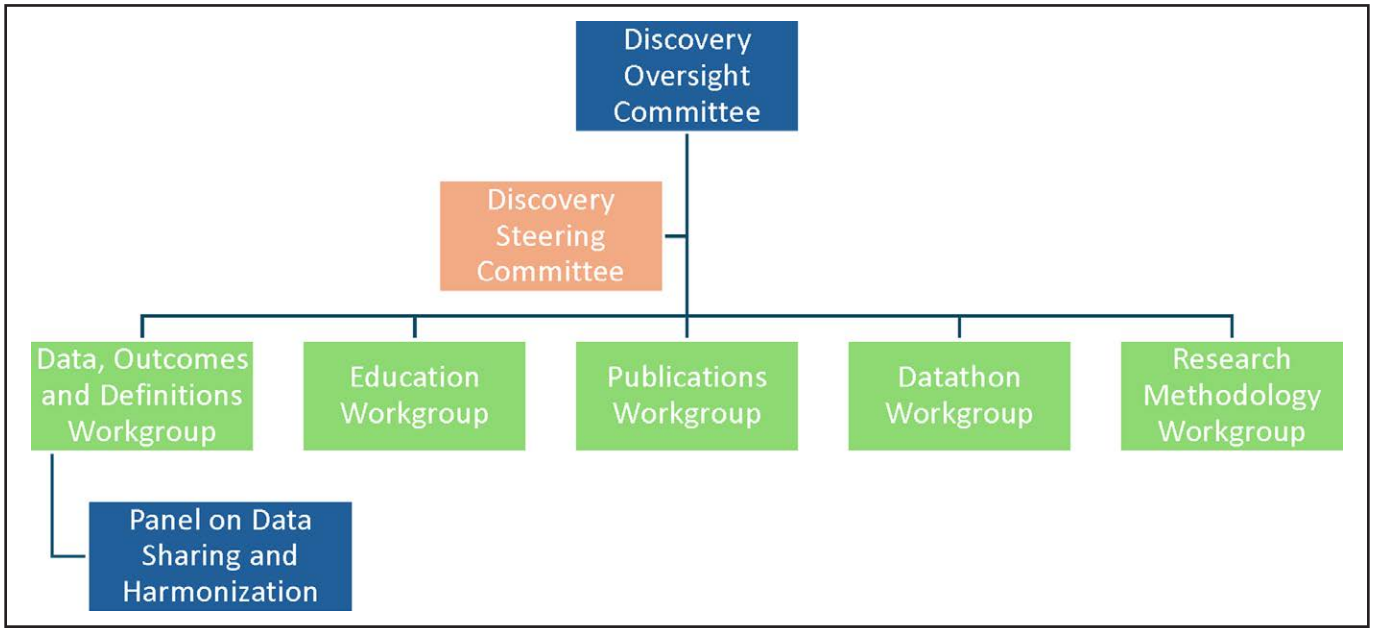

Figure 1. Organizational chart of Society of Critical Care Medicine Discovery and working groups of the Data Science Campaign. The Panel on Data Sharing and Harmonization formed to address needs identified by the Data, Outcomes, and Definitions Workgroup.

harmonization in critical care practice and research. The modified Delphi and guiding principles will be reported in full in a subsequent publication.

FINDINGS

The conceptual framework of organizing topics and consensus definitions as adopted by the PDSH is presented in **Table 1**.

To facilitate large, multicenter studies and research that examines the continuum of care (e.g., post-ICU outcomes), we need a process to transform data to conform with agreed-upon standard definitions so that researchers can confidently analyze data from multiple institutions together and support the deployment of advanced, validated machine learning tools. Common data elements (CDEs) are standardized concepts in data science organized into data domains, and the process of transforming data into such a standardized structure is "data harmonization." Harmonized data allows for the application of consistent "data quality" assessments and development of a shared understanding of data limitations that might affect analyses, and critically supports rigorous and reproducible analytics across sites and sources. A system of digital infrastructure designed to promote the adoption of CDEs, data harmonization processes, and meaningful data quality assessments is considered to be "interoperable." Facilitating access to data requires a coordinated process for "data access and sharing" by multiple stakeholders, often called a governance structure. Finally, each of these topics must be informed by a robust framework for "ethics and equity."

Data Domains and Common Data Elements

Data domains are groups of variables or concepts that are further refined into CDEs that help provide organization and structure to a data collection effort. Some domains are easily defined, such as demographics and anthropometrics, and the data elements included in such domains are readily agreed upon (e.g., age, height, weight) even if the specific definitions and coding of values are not (e.g., race and ethnicity). Other domains may include laboratory tests, clinical measurements, or comorbidities. Specific variables may prove challenging to organize into such domains due to overlap (e.g., Acute Physiology and Chronic Health Evaluation score) and variation in representation across EHR systems. There may also be ethical and epistemological implications in organizing other data elements into such domains, such as if pregnancy is considered a comorbidity. In the EHR, variables are subset into distinct data concepts that may be collected repeatedly (e.g., daily weights) and associated with date and time stamps. **Table 2** includes definitions and examples of these tiers of variable attributes developed to parse

TABLE 1.
Organizing Topics Identified by the Data Harmonization and Sharing Guiding Principles Panel With Consensus Definitions

Topic	Definition
Data domains and common data elements	Subject areas (domains), organizing concepts (variables), and specifically operationalized definitions (common data elements) with a fit-for-purpose degree of specificity
Data harmonization	The process of transforming and mapping data to conform with a standardized structure (i.e., data domains and common data elements) such as a common data model
Data quality	Measures and measurement of limitations of a dataset that might impact or affect analyses, including plausibility of values and conformance to standards
Interoperability and digital infrastructure	The systems that promote the capture and exchange of data and metadata and platforms that facilitate the adoption of common data elements, data harmonization processes, and meaningful data quality assessments
Data access, sharing, and governance	The rules, regulations, and processes dictating how stakeholders contribute and obtain data from different sources
Ethics and equity	Philosophical considerations and principles promoting fairness throughout all aspects of gathering, collecting, sharing, and using data as well as ensuring balanced and appropriate interpretation of any results and findings

TABLE 2.
The Tiers of Variable Attributes

Attribute	Definition	Example	
Data domain	A broad category of variables and data elements organized around a section of data elements	Vital signs	Anthropometrics and demographics
Subdomain	Smaller group of variables within a domain	Blood pressure	Demographics
Concept	Specific measurement or variable	Mean arterial pressure	Age
Common data element	Concept with defined framework for collection including temporality and plausibility rules	Lowest mean arterial pressure in first 24 hr of ICU admission measured by arterial line	Patient age in years at time of ICU admission

the categorizations and provide a shared terminology across DSC publications.

In preparation for data collection as part of a case report form (CRF) or registry, researchers need to add more specificity to a variable or concept and develop a CDE. The CDE provides a standardized definition of a given variable including the "identifier" or what is being measured, the "unit(s)" of measurement, other "elements" or aspects of the measurement (e.g., temporality of collection), and contextual factors or "metadata" (20) which help interpret the measurement (e.g., normal physiologic ranges, devices used, method of collection). These components are organized into micro-schemas, which help ensure a given measure is fully captured with any other information that might influence the CDE's meaning. For example, a study exploring metabolic syndrome in patients with acute COVID-19 infections faced a significant limitation around one of the key independent variables, triglyceride levels (TGLs), which can be falsely elevated in patients receiving propofol (21). The original CRF did not capture the temporal relationship between TGL and propofol administration. To conduct meaningful analysis, the study team had to collect additional information according to the CDE: Triglyceride measured in mg/dL collected before the first administration of propofol. This microschema is illustrated in **Figure 2**.

CDEs should build on vocabulary standards widely adopted on the basis of community consensus (e.g., Prescription Norm [RxNorm], Logical Observation Identifiers Names and Codes [LOINC]) (22, 23) or through regulation or policy (e.g., *International Classification of Diseases*, Current Procedural Terminology). In the use case above, triglyceride could be linked to the LOINC 2571-8 and propofol to the RxNorm notation 8782. Such standards are maintained and updated frequently through discussion forums, policy updates, and refinement through peer reviewed publication. While consensus is developing for CDE in many specific critical care conditions (e.g., stroke, trauma) (24, 25) and populations (e.g., pediatrics) (26), the critical care community has yet to coalesce robust standards for core ICU data common across the breadth of the field. SCCM Discovery, uniquely positioned to foster crucial conversations and debates, can only achieve consensus through sustained and meaningful engagement from its diverse stakeholders, including patients, advocacy groups, academic and hospital leaders, industry representatives, medical societies, government officials, researchers, and clinicians.

Data Harmonization

Data harmonization is the process of mapping data from disparate data sources to conform with a given data standard. Such standards are often called common data models (CDMs) and include both an overarching schema as well as value sets defining clinical data elements. Readers may be familiar with standards established by: 1) the Patient-Centered Outcomes Research Institute (PCORnet), a data model originally based on the Mini-Sentinel CDM that is designed for patient-centered clinical effectiveness research (27, 28); 2) Informatics for Integrating Biology and the Bedside (i2b2), a data model based on a star schema containing a central fact table made up by observations about a patient, supporting querying patient-level data based on

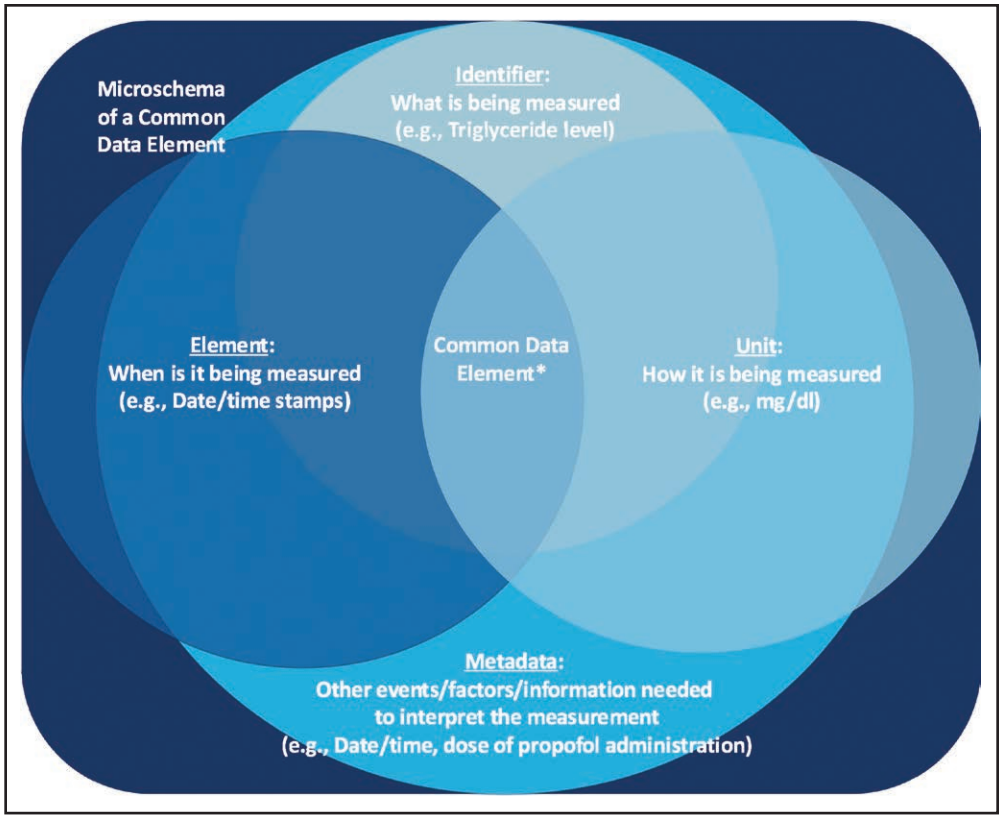

Figure 2. Example micro schema of a common data element (CDE) including components, aspects, and metadata of triglyceride measurement as captured in Denson et al (21). A plain-language definition of the CDE might be "Triglyceride as measured in mg/dL collected prior to the first administration of propofol."

specific criteria at local institutions (29, 30); or 3) the Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM), a data model built around the concept of observations collected about subjects who participated in a clinical study (31, 32). Pre-configured rules systems provide more robust definitions and ontologies defining the structure and relationship between data elements. For example, the Observational Medical Outcomes Partnership (OMOP) promotes harmonization while preserving the granularity of source EHR data (i.e., the microschema including metadata) (33, 34). These standards can be powerful tools for all data needs; however, processing data into this format takes more computational time, increasing data latency, or the time between generation of data and availability for secondary use.

Broadly, data harmonization occurs in three tiers: model alignment, value set mapping, and identifier mapping. Model alignment involves the mapping of schemas, or the interconnections of different tables in the dataset. A well-developed data model will support the contextualization of clinical events, treatments, and assessments. Value set mapping is the process of harmonizing concepts and elements from different schemas (e.g., race categories may be different across sources, have different labels, or may be derived from different terminologies). Identifier mapping involves term-to-term mapping to determine equivalency or relationships (35, 36). Researchers often merge multiple datasets or registries to provide larger or more diverse samples for analysis, but many studies still have significant limitations (**Table 3**); for example, multiple datasets may contain the same measure with different names and these may need to be reconciled through microschema and/or identifier mapping.

Researchers should understand that data harmonization standards are often designed for specific use cases. For example, CDISC SDTM is primarily designed for the representation of data from clinical trials for regulatory submission purposes, while the OMOP CDM is designed for the analysis of observational healthcare data from real-world sources. The optimal use of these standards depends on the nature of the data as well as the objectives of the research or analysis (37). Alternatively, the Fast Healthcare Interoperability Resources (FHIR) is an exchange standard that supports exchange of electronic healthcare data across different systems using community-specific specifications with so-called "implementation guides." When combined with a standard such as OMOP, both data science and application deployment research can be supported (**Fig. 3**) (38). A simple, albeit imperfect, metaphor could be the invention of the telephone. It is certainly a lot faster to share information with someone

TABLE 3.
Example Challenge in Harmonizing Data Elements Across Datasets

Variable Name	Definition	Source
DM	A disease in which the body does not control the amount of glucose (a type of sugar) in the blood and the kidneys make a large amount of urine. This disease occurs when the body does not make enough insulin or does not use it the way it should	Unified Medical Language System (23)
DM, with further classification as type 1 DM, type 2 DM, and gestational diabetes	Fasting plasma glucose ≥ 7.0 mmol/L (126mg/dL) or 2-hr plasma glucose ≥ 11.1 mmol/L (200 mg/dL) and determined according to electronic health records, physician notes, and ICD codes	Viral Infection and Respiratory Illness Universal Study COVID-19 Registry (24)
DM due to underlying condition without complications	E08.9 is a billable/specific ICD-10, Clinical Modification code	John Hopkins COVID-19 Precision Medicine Analytics Platform Registry (25)
Type 1 or type 2 DM	Requiring oral or subcutaneous treatment	International Severe Acute Respiratory and emerging Infection Consortium (26)
DM–Data Automation Definition	Hemoglobin A1c ≥ 6.5 and/or ICD-10 codes (e.g., E8, E9, E10, E11, E13, or E14) for DM	CURE ID (27)

DM = diabetes mellitus, ICD-10 = *International Classification of Diseases*, 10th Edition.

on the telephone (i.e., FHIR) than via telegram, but if one person speaks Mandarin (e.g., Epic) and the other speaks German (e.g., Cerner), the conversation will not be very productive. If we hire an interpreter (i.e., adopt a CDM or preconfigured rules system), the conversation may be slower, but we will finally understand each other. It is important to note, however, that an interpreter can introduce bias, so a robust data harmonization effort should preserve the original data and clear documentation of all transformations (i.e., data provenance) to ensure data users can assess for bias independently.

Throughout any harmonization effort, it is essential to maintain comprehensive documentation of provenance, a clear accounting of the source, transformation, and translation of a piece of data (35).

Even with the advanced terminology, vocabularies, and ontology captured in solutions like i2b2, PCORnet, and OMOP, significant gaps exist for critical care data. Data standards do not exist for many of the high-throughput data sources and a number of variables and data elements available in ICUs. Critical care experts must collaborate with data scientists to bridge this gap and provide a strong foundation of data standards to support development of CDE.

Data Quality

Assessing the quality of a given dataset often depends on the context. Quality is defined by the data's fit for its intended uses, while data integrity refers to the accuracy and consistency of the data over its life cycle. Inextricably linked with data quality is the notion of bias: systematic misrepresentations, mis-categorizations, or mis-interpretations of data or results that disadvantage certain groups of people (39). These may result from the perceptions of patients or clinicians or even from scientific literature. For instance, documented biases in healthcare results in Black women's pain being under-recognized, and race-based adjustments in kidney function tests disproportionately restrict kidney failure treatments for African American patients (40). The idea that data quality tests could or should eliminate all biases should be dismissed. Instead, clinicians and researchers should look for meaningful data quality assessments that fully articulate the limitations and potential biases of all datasets used in a given critical care context. Clinical domain experts must be vigilant and partner with data scientists to carefully define what constitutes bias and optimize data quality assessments in their specific analytical contexts. Conversely, data scientists and

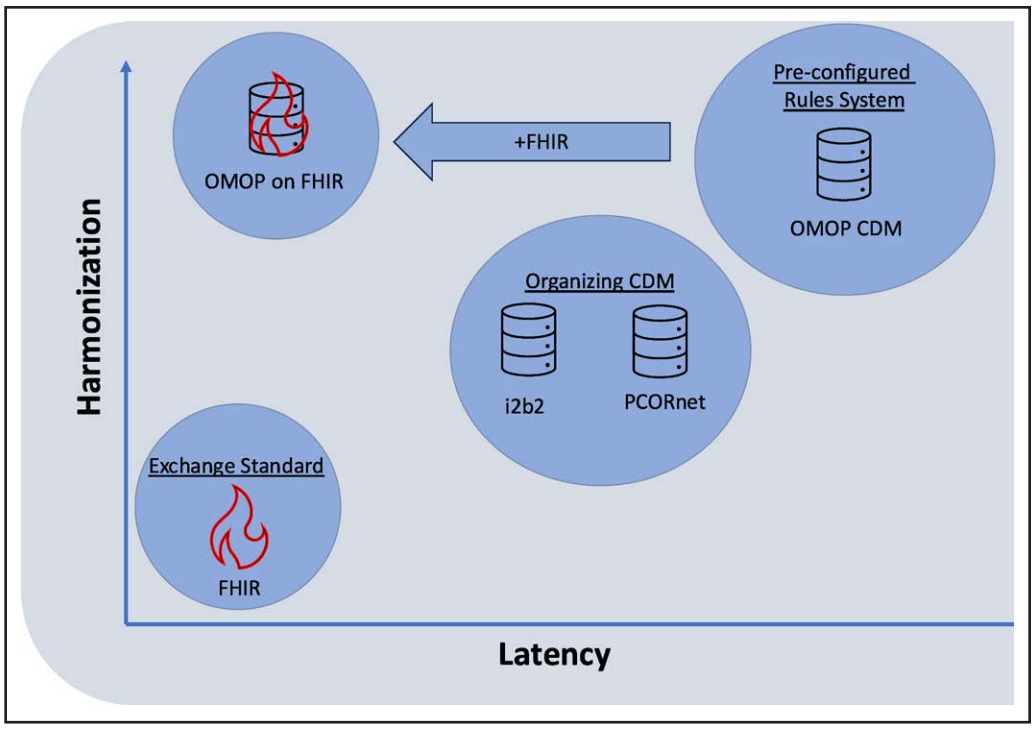

Figure 3. Both common data models (CDMs) and Fast Healthcare Interoperability Resources (FHIR) are necessary for interoperability and cross-site analytics. Patient-Centered Outcomes Research Institute (PCORnet), Informatics for Integrating Biology and the Bedside (i2b2) (both organizing CDMs), and the Observational Medical Outcomes Partnership (OMOP) CDM (a pre-configured rules system) can increase their interoperability and decrease latency through the use of FHIR standards while supporting both data science and application deployment research.

solution developers should be cognizant of the "messiness" of data in the real world and be careful not to develop models which are dependent on a ideal datasets (41–43).

Datasets must be reliable, dependable, and valid, meaning that the findings and insights gained from analysis can be trusted, are reproducible, and accurate through routine rules and constraints. Common measures of data quality also include conformance (e.g., does the data match the intended format), completeness (e.g., how frequently information is missing), and concordance (e.g., how well does the data agree across sources or expected parameters) (43, 44). A dataset must also be plausible, both temporally and atemporally (44). For example, a record indicating that a patient's height shrunk 6 inches over the first 48 hours could violate temporal plausibility rules, while an atemporal violation might exist for an adult's height documented as 10 inches. Furthermore, unit harmonization ensures data conforms to a canonical unit without sacrificing data integrity (41). A dataset must meet specified thresholds for both missingness and conformance to the model it is intended for, ensuring its suitability for the intended purpose.

Data Interoperability and Digital Infrastructure

Data interoperability is the degree to which information from one source and context can be understood and analyzed in another. Two studies may have a different definition of a variable, but the databases are interoperable if we can consistently and reliably translate or transform the data to be analyzed together (45). In other words, systems of data collection and storage are interoperable with each other when they each contain sufficient granularity to facilitate harmonization. Interoperability, then, is not a static attribute or accomplishment of a given dataset or source, but rather a metric deeply dependent on the context of each specific research question (46).

Interoperability can be improved through the adoption of CDE or harmonization efforts but should be considered separately (47). For example, blood pressure (BP) is an essential element/variable to include in almost any critical care study and may refer to any of three measurements (systolic BP, diastolic BP, or mean arterial pressure) and that different collection methods are available (e.g., manual auscultation, automated BP cuff, arterial line). Patient position and anatomy, equipment (e.g., size of cuff or arterial line patency), and a variety of clinical factors and medications influence these measurements. While a well-defined CDE could be adopted across all studies capturing BP, this may not always be possible, especially for studies that are already completed (47). In extracted EHR data, vocabularies and terminologies (e.g., LOINC) help capture most of the necessary data for BP, however,

the standardized vocabularies do not exist to harmonize data a priori in all cases, especially in emerging fields such as genomics, imaging, or waveform data. Interoperability is still possible when the system of data capture and storage facilitates exploration of sufficient granularity and nuance leveraging other interoperability approaches such as ontologies (10). Ultimately, the context of a study dictates the level of detail needed to ensure interoperability.

Data Access, Sharing, and Governance

Linked with digital infrastructure are the concepts of data access, sharing, and governance. High impact data is organized and curated to support appropriate use of said data for research, evaluation, and quality improvement efforts.

Data access is normally split into three tiers, which include fully identified, limited identification, and de-identified data (48–50). A fourth tier is sometimes provided—synthetic data. Such tiers help protect patient privacy and compliance with federal requirements (e.g., HIPAA and General Data Protection Regulation [GDPR]) by limiting access to sensitive protected health information (PHI) while providing comparable data access to support a wide range of analyses. Data access tiers need to be carefully designed and operationalized based on the specific subject matter requirements, Institutional Review Board oversight for identifiable or HIPAA-limited data, and consent for secondary data use. Researchers should avail themselves of expert opinions to determine the risk of re-identification, even when PHI is removed (51, 52).

Fundamentally, data governance is about who makes decisions for data access and how they are made. Many models exist, including those focused on who is providing the data or patient consent, but transparency and version control are essential to the success of any data sharing initiative.

To promote participation and collaboration, many large-scale data collection efforts include attribution and publication requirements. For instance, the SCCM Discovery Viral Infection and Respiratory Illness Universal Study COVID-19 registry required all ancillary studies to include co-authors from sites contributing the largest proportions of patients (53).

Ethics and Equity

Underpinning all other topics, careful attention must be paid to questions of ethics and equity. All data users must be aware of how clinical data can be (and has been) used to directly and indirectly harm patients, populations, and groups. Further, it would benefit researchers and patients to consider adopting a culture of transparency and extensively communicating the intended use of data (54, 55). Ethical stewardship of data demands transparency, informed consent, and safeguarding of patients' privacy, while equity necessitates the deliberate inclusion of diverse populations to ensure applicability of research findings across diverse demographics. Regulations including the HIPAA and GDPR provide broad frameworks for protecting patient privacy. However, incomplete alignment of these policies limits international data sharing, and heterogeneity in interpretation may leave patients vulnerable (54, 56, 57). Emerging technologies, specifically distributed ledgers (e.g., blockchain) offer the promise of new approaches to balancing the power dynamic between patients and data holders (58). While patients are generally supportive of their data being used for many purposes (12), greater transparency and improved oversight are clear priorities (59, 60). Guiding principles on the use of EHR data are imperative for protecting patients' rights and maintaining the scientific integrity of research and outcomes.

FUTURE DIRECTIONS

This primer for the critical care community serves as the foundation applying data science methodologies to datasets. These concepts help in extracting valuable insights from data, embodying a set of core principles guiding this process. Using this foundation, data mining techniques such as machine learning can be employed to extract knowledge from datasets. Additional emerging technologies can also be further assessed and validated by the community—from confidential computing modalities (e.g., zero knowledge proofs, homomorphic encryption), to federated learning and collaborative computing (i.e., data stays at rest and compute is sent to data) to blockchain and distributed ledger technologies for tracking data provenance and use (61). Furthermore, domain expertise plays a crucial role, particularly in critical care, as it entails a deep understanding of real-world clinical challenges and

patient care dynamics. This primer serves as the basis to contextualize the application of data science methodologies to healthcare issues, enhancing their efficacy and relevance.

Because it is the nature of scientific and clinical research to continuously evolve, standards must therefore also avoid stagnation. Furthermore, adherence to data standards is crucial for regulatory compliance, contributing to patient safety. Robust standards promote and facilitate data sharing and more rapid, larger scale research to advance clinical care, an essential step toward the creation of national clinical registries for critical care (62).

The rapid advancement of data science in critical care necessitates the creation and maintenance of living documents and guides promoting best practices for critical care researchers. As data science continues to evolve rapidly in critical care, there is a growing need for dynamic guidance that promotes best practices among clinicians and researchers. The living guidance document can rapidly adapt to the changing landscape of the critical care data science environment. Critical care clinicians are encouraged to actively engage in the SCCM Community to contribute and further evolve these standards. The SCCM's effort through its data science activities are a collective stakeholder effort to help shape and influence the future of data science in critical care, to ensure that the data standards effectively meet the diverse needs and perspectives within the field.

ACKNOWLEDGMENTS

We thank Chase Hamilton and Kevin Boyer, MPH from the Society of Critical Care Medicine for their support in preparation and review of this article.

1 *Critical Path Institute, Tucson, AZ.*

2 *Department of Public Health Sciences, Clemson University, Clemson, SC.*

3 *Society for Critical Care Medicine, Mount Prospect, IL.*

4 *Center for Advanced Analytics, Baptist Health South Florida, Miami, FL.*

5 *Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA.*

6 *Department of Mathematics, University of Pittsburgh, Pittsburgh, PA.*

7 *Critical Care Institute, Keck Hospital of USC, Los Angeles, CA.*

8 *Division of Trauma, Emergency Surgery and Surgical Critical Care, Department of Surgery, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA.*

9 *AI MINDSystems Foundation, Washington, DC.*

10 *Department of Pediatrics, Division of Pediatric Critical Care Medicine, UH Rainbow Babies and Children's Hospital, Case Western University School of Medicine, Cleveland, OH.*

11 *Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, University Hospital of Cleveland, Case Western University School of Medicine, Cleveland, OH.*

12 *School of Nursing, Duke University, Durham, NC.*

13 *School of Medicine, Duke University, Durham, NC.*

14 *School of Medicine, University of North Carolina, Chapel Hill, NC.*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccejournal).

This work was funded by the Society of Critical Care Medicine Discovery's Data Science Campaign.

Drs. Heavner, Kumar, and Anderson are co-investigators in the study Aggregating and Analyzing COVID-19 Treatments from Electronic Health Records and Registries Globally (CURE ID), funded by U.S. Food and Drug Administration (FDA) and Assistant Secretary for Planning and Evaluation (ASPE). Ms. Dasher's position was funded under CURE ID, funded by FDA and ASPE. Dr. Clermont is the Cheif Medical Officer and equity holder at NOMA AI. Dr. Cobb has an equity interest in Bauhealth, Aikido laboratories, and GibLib. Dr. Remy is a member of Immune Functional Diagnostics, LLC, which is unrelated to this work. Dr. Haendel is a founder of Alamya Health and a B-corp focused on genomic diagnostics. The remaining authors have disclosed that they do not have any potential conflicts of interest.

Society of Critical Care Medicine (SCCM) Discovery Panel on Data Sharing and Harmonization members are available at: http://links.lww.com/CCX/B436.

For information regarding this article, E-mail: sheavne@clemson.edu

REFERENCES

1. Toseef M, Olayemi Petinrin O, Wang F, et al: Deep transfer learning for clinical decision-making based on high-throughput data: Comprehensive survey with benchmark results. *Brief Bioinform* 2023; 24:bbad254
2. Yang S, Stansbury LG, Rock P, et al: Linking big data and prediction strategies: Tools, pitfalls, and lessons learned. *Crit Care Med* 2019; 47:840–848
3. Leese P, Anand A, Girvin A, et al: Clinical encounter heterogeneity and methods for resolving in networked EHR data: A study from N3C and RECOVER programs. *J Am Med Inform Assoc* 2023; 30:1125–1136
4. Kohane IS, Aronow BJ, Avillach P, et al; Consortium For Clinical Characterization Of COVID-19 By EHR (4CE): What every

reader should know about studies using electronic health record data but may be afraid to ask. *J Med Internet Res* 2021; 23:e22219

5. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016; 3:160018
6. Kerstenetzky L, Birschbach MJ, Beach KF, et al: Improving medication information transfer between hospitals, skilled-nursing facilities, and long-term-care pharmacies for hospital discharge transitions of care: A targeted needs assessment using the Intervention Mapping framework. *Res Social Adm Pharm* 2018; 14:138–145
7. Gamal A, Barakat S, Rezk A: Standardized electronic health record data modeling and persistence: A comparative review. *J Biomed Inform* 2021; 114:103670
8. Healthcare Information and Management Systems Society: Sharing Data Across the Care Continuum: Perspectives From Long-Term and Post-Acute Care Case Studies. 2021. Available at: https://gkc.himss.org/resources/sharing-data-across-care-continuum-perspectives-long-term-and-post-acute-care. Accessed August 11, 2024
9. Pylypchuk Y, Johnson C, Patel V: State of Interoperability Among U.S. Non-Federal Acute Care Hospitals in 2018. Office of the National Coordinator for Health Information Technology. 2020. Available at: https://www.healthit.gov/data/data-briefs/state-interoperability-among-us-non-federal-acute-care-hospitals-2018. Accessed August 11, 2024
10. Palojoki S, Lehtonen L, Vuokko R: Semantic interoperability of electronic health records: Systematic review of alternative approaches for enhancing patient information availability. *JMIR Med Inform* 2024; 12:e53535
11. Hohman KH, Martinez AK, Klompas M, et al: Leveraging electronic health record data for timely chronic disease surveillance: The multi-state EHR-based network for disease surveillance. *J Public Health Manag Pract* 2023; 29:162–173
12. Cascini F, Pantovic A, Al-Ajlouni YA, et al: Health data sharing attitudes towards primary and secondary use of data: A systematic review. *EClinicalMed* 2024; 71:102551
13. Richter G, Borzikowsky C, Hoyer BF, et al: Secondary research use of personal medical data: Patient attitudes towards data donation. *BMC Med Ethics* 2021; 22:164
14. Stetler CB, Legro MW, Wallace CM, et al: The role of formative evaluation in implementation research and the QUERI experience. *J Gen Intern Med* 2006; 21:S1–S8
15. Elwy AR, Wasan AD, Gillman AG, et al: Using formative evaluation methods to improve clinical implementation efforts: Description and an example. *Psychiatry Res* 2020; 283:112532
16. Palinkas LA, Cooper BR: Mixed methods evaluation in dissemination and implementation science. In: *Dissemination and Implementation Research in Health: Translating Science to Practice*. Second Edition. Brownson RC, Colditz GA, Proctor EK (Eds). New York, Oxford Academic, 2017. Available at: https://doi.org/10.1093/oso/9780190683214.003.0020. Accessed August 11, 2024
17. McNall M, Foster-Fishman PG: Methods of rapid evaluation, assessment, and appraisal. *Am J Eval* 2007; 28:151–168
18. Vindrola-Padros C, Brage E, Johnson GA: Rapid, responsive, and relevant?: A systematic review of rapid evaluations in health care. *Am J Eval* 2021; 42:13–27
19. McKim C: Meaningful member-checking: A structured approach to member-checking. *Am J Qual Res* 2023; 7:41–52
20. Ulrich H, Kock-Schoppenhauer AK, Deppenwiese N, et al: Understanding the nature of metadata: Systematic review. *J Med Internet Res* 2022; 24:e25440
21. Denson JL, Gillet AS, Zu Y, et al; Society of Critical Care Medicine Discovery Viral Infection and Respiratory Illness Universal Study (VIRUS): COVID-19 Registry Investigator Group: Metabolic syndrome and acute respiratory distress syndrome in hospitalized patients with COVID-19. *JAMA Netw Open* 2021; 4:e2140568
22. National Library of Medicine: RxNorm. Available at: https://www.nlm.nih.gov/research/umls/rxnorm/index.html. Accessed June 1, 2024
23. Logical Observation Identifiers Names and Codes. Available at: https://loinc.org/. Accessed June 1, 2024
24. Common Data Elements: Stroke. Available at: https://www.commondataelements.ninds.nih.gov/Stroke. Accessed June 1, 2024
25. American College of Surgeons: National Trauma Data Standard. Available at: https://www.facs.org/quality-programs/trauma/quality/national-trauma-data-bank/national-trauma-data-standard/. Accessed June 1, 2024
26. Ward SL, Flori HR, Bennett TD, et al: Design and rationale for common data elements for clinical research in pediatric critical care medicine. *Pediatr Crit Care Med* 2020; 21:e1038–e1041
27. The National Patient-Centered Clinical Research Network. Available at: https://pcornet.org/. Accessed June 1, 2024
28. Wagholikar KB, Jain R, Oliveira E, et al: Evolving research data sharing networks to clinical app sharing networks. *AMIA Jt Summits Transl Sci Proc* 2017; 2017:302–307
29. Informatics for Integrating Biology and the Bedside. Available at: https://www.i2b2.org/. Accessed June 1, 2024
30. Post AR, Pai AK, Willard R, et al: Metadata-driven clinical data loading into i2b2 for Clinical and Translational Science Institutes. *AMIA Jt Summits Transl Sci Proc* 2016; 2016:184–193
31. Clinical Data Interchange Standards Consortium: Study Data Tabulation Model. Available at: https://www.cdisc.org/standards/foundational/sdtm. Accessed June 1, 2024
32. Wood F: The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics. Proceedings of the Pharmaceutical Industry SAS® Users Group, Atlanta, GA, June 1-4, 2008. Available at: http://www.lexjansen.com/pharmasug/2008/rs/RS10.pdf. Accessed August 11, 2024
33. Observational Health Data Sciences and Informatics. Available at: https://www.ohdsi.org/. Accessed June 1, 2024
34. Park K, Cho M, Song M, et al: Exploring the potential of OMOP common data model for process mining in healthcare. *PLoS One* 2023; 18:e0279641
35. Matentzoglu N, Balhoff JP, Bello SM, et al: A Simple Standard for Sharing Ontological Mappings (SSSOM). *Database* 2022; 2022:baac035
36. McMurry JA, Juty N, Blomberg N, et al: Identifiers for the 21st century: How to design, provision, and reuse persistent

identifiers to maximize utility and impact of life science data. *PLoS Biol* 2017; 15:e2001414

37. Schneeweiss S, Brown JS, Bate A, et al: Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products. *Clin Pharmacol Therap* 2020; 107:827–833
38. Ayaz M, Pasha MF, Alzahrani MY, et al: The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities. *JMIR Med Inform* 2021; 9:e21929
39. Ranard BL, Park S, Jia Y, et al: Minimizing bias when using artificial intelligence in critical care medicine. *J Crit Care* 2024; 82:154796
40. Levey AS, Titan SM, Powe NR, et al: Kidney disease, race, and GFR estimation. *Clin J Am Soc Nephrol* 2020; 15:1203–1212
41. Pfaff ER, Girvin AT, Gabriel DL, et al: Synergies between centralized and federated approaches to data quality: A report from the national COVID cohort collaborative. *J Am Med Inform Assoc* 2021; 29:609–618
42. Cook L, Espinoza J, Weiskopf NG, et al; N3C Consortium: Issues with variability in electronic health record data about race and ethnicity: Descriptive analysis of the National COVID Cohort Collaborative Data Enclave. *JMIR Med Inform* 2022; 10:e39235
43. Lewis AE, Weiskopf N, Abrams ZB, et al: Electronic health record data quality assessment and tools: A systematic review. *J Am Med Infor Assoc* 2023; 30:1730–1740
44. Kahn MG, Callahan TJ, Barnard J, et al: A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016; 4:1244
45. Canova-Barrios C, Machuca-Contreras F: Interoperability standards in health information systems: Systematic review. *Sem Med Writ Educat* 2022; 1:7
46. Bernstam EV, Warner JL, Krauss JC, et al: Quantitating and assessing interoperability between electronic health records. *J Am Med Inform Assoc* 2022; 29:753–760
47. Kush RD, Warzel D, Kush MA, et al: FAIR data sharing: The roles of common data elements and harmonization. *J Biomed Inform* 2020; 107:103421
48. Suver C, Harper J, Loomba J, et al: The N3C governance ecosystem: A model socio-technical partnership for the future of collaborative analytics at scale. *J Clin Transl Sci* 2023; 7:e252
49. Suver C, Thorogood A, Doerr M, et al: Bringing code to data: Do not forget governance. *J Med Internet Res* 2020; 22:e18087
50. Mayo KR, Basford MA, Carroll RJ, et al: The all of us data and research center: Creating a secure, scalable, and sustainable ecosystem for biomedical research. *Ann Rev Biomed Data Sci* 2023; 6:443–464
51. Xia W, Liu Y, Wan Z, et al: Enabling realistic health data re-identification risk assessment through adversarial modeling. *J Am Med Inform Assoc* 2021; 28:744–752
52. Zhang X, Wan Z, Yan C, et al: How adversarial assumptions influence re-identification risk measures: A COVID-19 case study. In: *Privacy in Statistical Databases*. Domingo-Ferrer J, Laurent M (Eds). Cham, Switzerland, Springer International Publishing, 2022, pp 361–374
53. Society of Critical Care Medicine: Discovery VIRUS COVID-19 Registry. Available at: https://www.sccm.org/Research/Research/Discovery-Research-Network/VIRUS-COVID-19-Registry. Accessed June 1, 2024
54. Sanchez-Pinto LN, Luo Y, Churpek MM: Big data and data science in critical care. *Chest* 2018; 154:1239–1248
55. Perrino T, Howe G, Sperling A, et al; NIMH Collaborative Data Synthesis for Adolescent Depression Trials Study Team including: Advancing science through collaborative data sharing and synthesis. *Perspect Psychol Sci* 2013; 8:433–444
56. U.S. Department of Health and Human Services: Common Rule. Available at: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html. Accessed August 11, 2024
57. de Kok J, de la Hoz MAA, de Jong Y, et al: A guide to sharing open healthcare data under the General Data Protection Regulation. *Sci Data* 2023; 10:404
58. Gross M, Miller R: Protecting Privacy and Promoting Learning: Blockchain and Privacy Preserving Technology Should Inform New Ethical Guidelines for Health Data. 2019
59. Kassam I, Ilkina D, Kemp J, et al: Patient perspectives and preferences for consent in the digital health context: State-of-the-art literature review. *J Med Internet Res* 2023; 25:e42507
60. Canaway R, Boyle DI, Manski-Nankervis JE, et al: Gathering data for decisions: Best practice use of primary care electronic records for research. *Med J Aust* 2019; 210:S12–S16
61. Passerat-Palmbach J, Farnan T, McCoy M, et al: Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data. *IEEE International Conference on Blockchain (Blockchain).* 2020, pp 550–555
62. Salluh JIF, Quintairos A, Dongelmans DA, et al; Linking of Global Intensive Care (LOGIC) and Japanese Intensive care PAtient Database (JIPAD) Working Group: National ICU registries as enablers of clinical research and quality improvement. *Crit Care Med* 2024; 52:125–135